Introduction to Artificial Intelligence, Fall & Winter 2022
College of Computer Science, Zhejiang University
Reference Solutions for Problem Set 4

丁尧相

2023 年 1 月 6 日

Problem 1.(Logistic regression) 在 Lecture 11 幻灯片的第 15 页，我们给出了使用对率线性概率模型时，二分类 logistic loss 的一种形式：

$$L(\mathbf{x}, y; f) = \log(1 + e^{-y\mathbf{w}^T\mathbf{x}}), \quad y \in \{-1, 1\}. \tag{1}$$

而在第 17 页，我们则给出了同样使用对率线性概率模型时，一般 C 分类（$C \geq 2$）下的 logistic loss：

$$L(x, y; f) = -\sum_{c=1}^{C} \mathbb{I}[y = y_c] \log p(y_c|\mathbf{x}), \quad p(y_c|\mathbf{x}) = \frac{e^{\mathbf{w}_c^T\mathbf{x}}}{\sum_{i=1} e^{\mathbf{w}_i^T\mathbf{x}}}. \tag{2}$$

试论证公式（1）中定义的二分类损失，和公式（2）在 $C = 2$ 时的损失本质上是等价的。

参考解答: 对于公式（2），当正确标记 $y = +1$ 时，损失形式为

$$L(x, y; f) = \log(e^{(\mathbf{w}_2 - \mathbf{w}_1)^T\mathbf{x}} + 1).$$

当正确标记 $y = -1$ 时，损失形式为

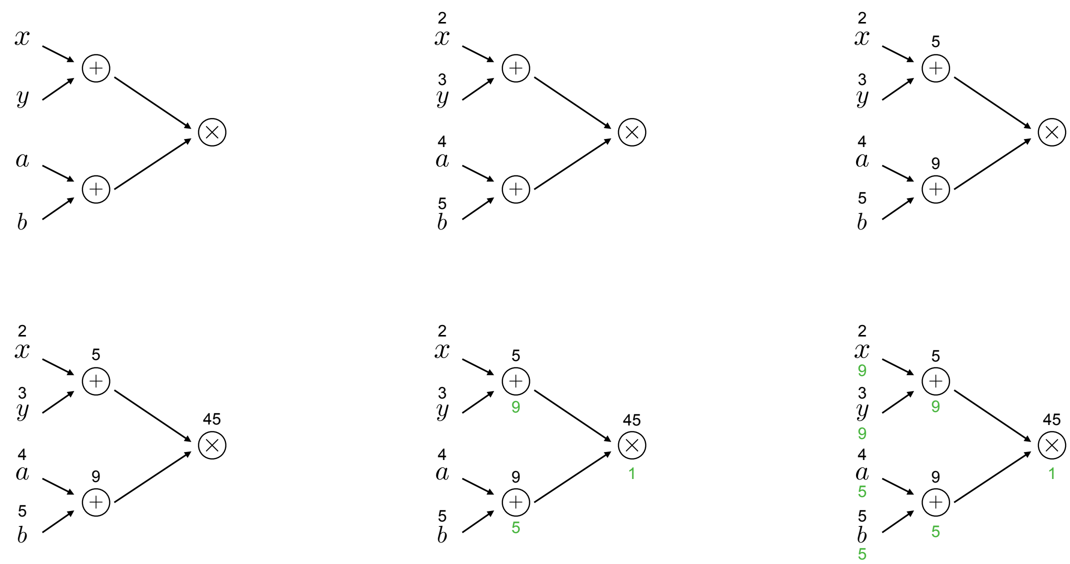
$$L(x, y; f) = \log(e^{(\mathbf{w}_1 - \mathbf{w}_2)^T\mathbf{x}} + 1).$$

可见公式（1）中的 $\mathbf{w}$ 可以代换为公式（2）中的 $\mathbf{w}_1 - \mathbf{w}_2$。

Problem 2.(Back propagation) 请参考 Lecture 11 幻灯片第 48 页，回答下面的问题：

- 画出函数 $f(x, y, a, b) = (x + y)(a + b)$ 所对应的计算图。
- 给定 $x = 2, y = 3, a = 4, b = 5$，请根据计算图给出利用反向传播算法计算各变量偏导数值的过程。

参考解答: 如图所示。第一图为计算图。后续图为 BP 算法运行流程，变量上方黑色数值为前向运行函数值，变量下方绿色数值为梯度值。



Problem 3.(Convolution) 设输入图像大小为 $84 \times 84 \times 3$。请设计两个不同的具有三个卷积（可带 pooling）层的网络结构，即指定每一层卷积的 kernel size, filter 个数，有无 padding，stride 大小，有无 pooling 及 pooling 尺寸，使得通过这三个卷积层后，输出尺寸为 $12 \times 12 \times 64$。

参考解答: 下图给出两个例子。

	Layer1	Layer2	Layer3
kernel size	4*4	4*4	3*3
filter数	16	32	64
padding	有（各边+1行/列）	有（各边+1行/列）	有（各边+2）
stride	2	2	2
pooling	无	无	无

	Layer1	Layer2	Layer3
kernel size	4*4	3*3	3*3
filter数	32	64	64
padding	无	有（各边+2行/列）	有（各边+1）
stride	4	2	1
pooling	无	无	无

Problem 4.(***)(Activation function) 在 Lecture 12 中，我们讲到 gradient vanishing 是妨碍神经网络层数变深的主要障碍之一，而使用更合适的激活函数可以有效缓解这一问题，据此，请回答下面的问题:

- 请说明为什么采用 Relu: $f(x) = \max(0, x)$ 作为激活函数，比采用 Sigmoid: $f(x) = \frac{1}{1+e^{-x}}$ 能更好

地缓解梯度消失问题。

- 假设在 MLP 网络中使用 Sigmoid 作为激活函数。设该神经元激活前的函数表达式为 $g = -\mathbf{w}^T Sig(g') + b$，其中 Sig 即为 Sigmoid 激活函数，$\mathbf{w}, b$ 分别为权重向量和偏置，g' 为上一层神经元激活前的输出向量。试证明 $\mathbf{w}$ 每一维对应的偏导数总是同号的。
- 上一问的现象对神经网络的学习有没有负面影响？为什么？
- 若改为使用 Relu 作为激活函数，是否存在上面的问题？

参考解答: 请参考https://web.eecs.umich.edu/ justincj/slides/eecs498/FA2020/598_FA2020_lecture10.pdf第 7-35 页内容。